Integrating Computer Assisted Learning into a Regular Curriculum: Evidence from a Randomized Experiment in Rural Schools in Shaanxi

Di Mo[1]
Stanford University
LICOS Centre for Institutions and Economic Performance, KU Leuven
CCAP, IGSNRR, Chinese Academy of Sciences

Linxiu Zhang[2], Renfu Luo[3] and Qinghe Qu[4]
CCAP, IGSNRR, Chinese Academy of Sciences

Weiming Huang[5],
Colby College

Jiafu Wang[6]
Indiana University Bloomington

Yajie Qiao[7]
Ankang University

Matthew Boswell[8] and Scott Rozelle[9]
Stanford University

Abstract:

Recent attention has been placed on whether computer assisted learning (CAL) can effectively improve learning outcomes. However, the empirical evidence of its impact is mixed. Previous studies suggest that the lack of an impact in developed countries may be attributable to substitution of effort/time away from productive, in-school activities. However, there is little empirical evidence on how effective an in-school program may be in developing countries. In order to explore the impact of an in-school CAL program, we conducted a clustered randomized experiment involving over 4000 third and fifth grade students in 72 rural schools in China. Our results indicate that the in-school CAL program has significantly improved the overall math scores by 0.16 standard deviations. Both the third graders and the fifth graders benefited from the program.

[1] Email: dimo@stanford.edu. Phone: +1(650)3363621.
[2] lxzhang.ccap@igsnrr.ac.cn
[3] luorf.ccap@igsnrr.ac.cn
[4] quqinghe@yahoo.com.cn
[5] wmhuang89@gmail.com
[6] jiafwang@indiana.edu
[7] careline49@163.com
[8] kefka@stanford.edu
[9] rozelle@stanford.edu

Integrating Computer Assisted Learning into a Regular Curriculum: Evidence from a Randomized Experiment in Rural Schools in Shaanxi

Introduction

In the last decade, attention has been placed on initiatives that adopt computer technology to confront the long-standing challenge of delivering quality education to poor and disadvantaged populations (e.g., Banerjee et al. 2007; Barrow, 2008; Linden, 2008; Cristia et al., 2012; Guimarães et al., 2013). These studies aim to evaluate whether educational input, such as computer assisted learning (CAL) programs, can improve student learning. CAL programs utilize computers and modern computing technologies (including both software and hardware devices) to enhance learning through computerized instruction, drills and exercises (Kirkpatrick and Cuban, 1998; Present's Committee of Advisors on Science and Technology, 1997). When integrated with well-designed educational games, CAL programs can sustain the interest and curiosity of students and lead to gains in student performance. The program also allows for the delivery of a consistent curriculum regardless of training or expertise of the teachers (Nara and Noda, 2003). In some contexts, such a CAL program may be more cost-effective than using teachers to provide additional instruction (Banerjee et al, 2008).

Despite the popularity of investment in computer technology in education, the empirical evidence on the impact of such programs has been mixed. Early studies in Israel and the United States found little consistent evidence as to whether the application of computer technology in school instruction has beneficial effects for student academic achievement (e.g. Angrist and Lavy, 2002; Fuchs and Woosmann, 2004; Goolsbee and Guryan, 2006). Later studies utilizing randomized experiments to evaluate specific CAL programs also found mixed evidence. For example, both Dynarski et al. (2007) and Krueger and Rouse (2004) found no significant gain in math and reading test scores from CAL programs for students in the United States. In contrast, Barrow, Markman and Rouse (2008) found a CAL program improved student math test scores by 0.17 standard deviations in urban schools in three districts in the US. This particular CAL program used computer-aided instructions in algebra to replace traditional classroom teaching. Although relatively few in number, evaluations conducted in developing countries mostly show CAL has had positive effects on student test scores (Banerjee et al., 2007; He, Linden and MacLeod, 2008; Linden, 2008).

An important limitation to these studies is that they usually examine CAL as an educational input, and do not consider whether the program is rolled out as an *in-school program* or *out-of-school program*. Miettinen (1999) defines in-school programs as those that occur during regular school hours and/or those that are organized as formal classroom activities. In contrast, out-of-school programs take place after school hours and/or are built around less formal group activities.

Out-of-school programs have several potential advantages over in-school programs. First, out-of-school programs can be run without the restrictions of the formal classroom (e.g., the limited scope of activities that are allowed to take place under the supervision of formal teachers). Second, out-of-school programs can be run without taking time away from other regularly scheduled classroom learning activities. For these reasons, remedial education camps and other after school programs have been found to lead to higher levels of learning (Hull and Schultz, 2002). Third, out-of-school programs are also potentially easier to design in a way that effectively caters to the needs of individuals, as has been shown with professional training programs that teach workplace skills (Dias et al., 1999).

Despite their potential benefits, out-of-school programs also have several disadvantages. First, successfully implementing such programs often requires schools to make a variety of structural changes regarding curriculum, staff allocation and meal programs. Second, the success of out-of-school programs often depends on the extraordinary effort of teachers or volunteers. As a consequence, it is often suggested that for programs to be sustainable, they ultimately need to be incorporated into the regular school-day curriculum (Underwood et al., 2000). Furthermore, the advantages of out-of-school programs may disappear if in-school programs are organized to use educational resources more efficiently (Cole, 1996).

Previous research shows that an in-school CAL program is ineffective in improving student learning in India (Linden, 2008). In contrast, the same research team showed that an out-of-school version of the same program did improve student learning. According to Linden (2008), one of the reasons for the difference in impact is that the in-school program appears to have created a substitution of effort/time away from other productive, in-class activities (such as formal instruction by the teacher in math and language). The out-of-school program instead appears to have served as a complement to existing resources.

Despite the interesting findings, the Linden study did not take into account how the in-school CAL program was incorporated into the regular school curriculum. India is known for short school days. Schools often provide only five to six hours per day of instruction (as reported in the study by Duflo et al., 2008). Incorporating CAL during school hours may have to replace relatively productive teaching periods. There is also high absenteeism of teachers during regular school hours (Chaudhury, et al., 2005), which may have contributed to the ineffectiveness of the in-school CAL program, particularly if the supervising teachers frequently missed the CAL sessions.

In contrast, rural schools in China have many features that potentially make an in-school CAL program more effective. In rural China, a school day typically runs from seven to eight hours with an additional one-hour noon break. China is reported to have much lower teacher absenteeism than India (Liu and Kumar, 2008). Moreover, all Chinese schools are required to allocate time for computer, art or music classes. However, since many rural schools do not have teachers for these subjects, rural schools typically have multiple time slots in a week that are relatively unproductive. In other words, there may be less of a substitution effect if CAL programs are run in-school. Thus it is likely that an in-school CAL program in rural China may be more effective in improving student learning than a similar in-school program in India.

Previous studies that employed large-scale randomized controlled trials to evaluate

CAL interventions were all run as out-of-school programs (Lai et al., 2011; Mo et al., 2012; Lai et al., 2012; Lai et al., 2013). These studies have shown that CAL can significantly improve the math and Chinese test scores of rural students in China. However, in these previous efforts to test CAL, there was no danger of the CAL program substituting for other classroom activities run by teachers during the regular teaching day. The question remains whether a CAL program in Chinese schools will be equally effective in improving learning when it is implemented as an in-school program during regular school hours.

The overall goal of this paper is to explore the impact of an in-school CAL program on the academic outcomes of an underserved student population in a developing country. To achieve this goal, the main question that we seek to answer in this paper is whether an in-school CAL program increases school performance. To do this, we also address the questions of whether in-school CAL programs increase the academic performance of grade 3 students and grade 5 students, and how the in-school CAL effect compares with that of a CAL program implemented outside of the classroom.

The rest of the paper is organized as follows. The next section reviews the study by Lai et al. (2013), which highlights the importance of evaluation an in-school CAL program in light of large expected investments in computing infrastructure in China's rural schools. The third section describes the current study's methodology, including the research design and sampling, intervention design, data collection and statistical approach.

The remaining sections present the results from the study, discuss the findings and conclude.

In-school CAL Program in China

The question of whether an in-school CAL can improve student learning is of particular relevance to China. As part of a new effort to improve the facilities in rural schools, the government has recently invested in improving the computing infrastructure of rural public schools (Yuan, 2012). By 2011, 86 percent of the rural public schools had set up computer rooms with an average of 17 computers in each school (Yang et al., 2012). China's Ministry of Education, however, has even more ambitious plans. The recently announced 12th Five-Year Plan for Integrating Information Technology into Education aspires to set up a computer room in every rural school by 2020 (Ministry of Education, 2012). Since the plan requires such an enormous investment of fiscal resources, it is important to learn how to effectively use the new computing resources.

Unfortunately, China's rural schools face many constraints in providing quality computer-based education. Teachers at rural schools typically do not have the qualifications or materials necessary to promote learning in computer classes (Lai et al., 2011). Teachers lack the training and/or motivation to adequately instruct students and pique the interest of students in using computers for learning. There is also a shortage of curricula to use during computing class, particularly in poor rural areas (Yang et al., 2012). Although 69 percent of students in rural public schools have computer classes, research shows that few schools have employed computers and/or educational software for instructional purposes in core academic subjects. Even when they do, computer classes are frequently cancelled due to a lack of teachers and instructional materials.

If CAL classes could occur in periods already assigned for computer classes—which are not being used effectively—it may be that in-school CAL classes in China could bring the proven benefits of the CAL program without any offsetting effects. In other words, if CAL could be conducted during the regularly-scheduled but poorly-utilized computer class period, CAL in rural China may be both integrative and supplemental.

Sampling, Data and Methods

Sampling and the Process of Randomization

We conducted a clustered (at the school level) Randomized Controlled Trial of CAL in Shaanxi rural schools during the 2011-2012 academic school year. A total of 5267 students in 72 rural Shaanxi schools were involved in the study. The study covered third grade and fifth grade students.[1]

Choosing the sample consisted of several steps. First, to focus our study on students from poor rural areas, we restricted our sample frame to four counties randomly selected out of the ten counties in Ankang Prefecture, the prefecture that covers one of the poorest areas in the southern region of Shaanxi Province. Shaanxi Province is situated in northwest China, which is one of the poorest regions in the country (Ezroj et al., 2004). Shaanxi ranks the second place among all provinces in China (CNBS, 2013) in terms of number of nationally designated poor counties. Ankang prefecture (where our sample counties are selected from) covers one of the poorest areas in the southern part of Shaanxi

Province. Eighty percent of the counties in Ankang are nationally designated counties. The average per capita income of the randomly selected four counties was about 4000RMB ($650) per year in 2011, which is far below rural China's average per capita income of 6977RMB in the same year (CNBS, 2011). Three out of the four sample counties are nationally designated poor counties in China.[2]

After choosing the counties, we obtained a comprehensive list of all *wanxiao* (those elementary schools with six full grades—grade one through grade six) in each of the four counties from the Department of Education of Ankang Prefecture.[3] We included all 72 schools that met the above criterion in our sample.

Within the sample schools, we included both third grade and fifth grade students in the 72 schools in our sample. We chose third grade and fifth grade students for two reasons. First, at the time of the launch of the project, we only had remedial tutoring material for students from third to sixth grade. It is for this reason that we did not choose students from first grade or second grade. Second, a subset of the fourth grade and sixth grade students in the school had already participated in a pilot project during the previous academic year. In order to avoid confounding the treatment effect, we chose to focus the intervention on third grade and fifth grade students. Again, none of these students had ever participated in a CAL program prior to the 2011-12 academic year.

All of the third grade and fifth grade students in the 72 sample schools were included. In phase one of this study (Lai et al., 2013), we had only included students who boarded at school. In this study we included students who were boarding at school and students who were living at home. Of the total number of students involved in the study (5267), 2279 were third grade students and 2988 were fifth grade students (Figure 1).

Although at the time of the baseline survey the main sample included a total of 72 schools and 5267 students, for various reasons (mainly because of school transfers and extended absences due to illness or injuries), there was some attrition by the end of the study. By the time of the evaluation survey we were able to follow up with 4757 students in the 72 sample schools (Figure 1, final row). In other words, 4757 out of the initial 5267 students (who took the baseline survey) were included in our evaluation survey and were part of the subsequent statistical analysis; 9.8 percent of the sample dropped out between the baseline and endline surveys. There were 249 attrited students (10.9 percent) from the third grade and 261 attrited students (8.7 percent) from the fifth grade. Fortunately for the study's integrity, there were no variables that were systematically related between the characteristics of students and their attrition status (Table 1).

After choosing the 72 schools for our sample, we randomly assigned them to either the treatment or control group. This assignment was done after the baseline. During the baseline, both the enumerators and the respondents/participants were blind to their eventual group assignment. In order to assure that the treatment and control groups were similar in terms of key characteristics at the time of the baseline, we pre-balanced along several key variables when we randomized. These key variables include the control variables listed in Appendix 1 (i.e. *student gender, student age, boarding student, ever repeated a grade, only child*, *age of father, age of mother, father has at least junior high* *school degree, mother has at least junior high school degree, at least one parent lives at home and family wealth*). This method is discussed by Bruhn (2008). In doing so, we re-randomized several times until the key baseline variables that are listed in Appendix 1 in the revised manuscript were balanced between the treatment and the control groups.

After the randomization, 36 schools were assigned to receive the CAL intervention. As the CAL intervention engaged both third grade and fifth grade students, the 2435 students of the third and fifth grades in the 36 treatment schools constitute the treatment group (Figure 2). Among these students, there were 1067 third grade students and 1368 fifth grade students. The 2832 students (1212 from the third grade and 1620 from the fifth grade) in the other 36 schools served as the control group. Due to the attrition, there were 4757 students left in our final analytic sample, among whom 2220 were in the 36 treatment schools, and 2537 were in the control schools.

Experiment Arm/Intervention

The main intervention involved computer assisted math remedial tutoring sessions that were designed to complement the regular in-class math curriculum for the entire school year 2011-2012. Under the monitoring of two teacher-supervisors trained by our research group, the students in the treatment group had two 40-minute CAL sessions per week as regular classes in school.[45] The sessions were mandatory and attendance was taken by the teacher-supervisors.

According to our protocol, the CAL sessions were supposed to be given during the normal "computer class" time period. We chose the computer class time periods since typically these are reserved for teaching non-academic material. Based on our surveys, in the computer classes offered in most of China's rural schools, students were taught basic computer operations, such as using a mouse, typing Chinese and using Microsoft Office software. On average, in 75.6 percent of the rural public schools in Shaanxi Province students are taught such basic computer operations in computer classes. When the schools do not have computer teachers to teach the class, computer class time is frequently used for students to practice math, Chinese or English questions under teacher supervision.

The instructional videos and games that comprise the content of each CAL session were designed for improving students' basic competencies in the uniform national math curriculum. The content was exactly the same for all students within the same grade among schools in the treatment group. During each session, two students shared one computer and played math games designed to help students review and practice the basic math material that was being taught in their regular school math classes. In a typical session, the students first watched an animated video that reviewed the material that they were receiving instruction on during their regular math class sessions in that week. The students then played math games with animated characters to practice the skills introduced in the video lecture.[6] If students had a math-related question, they were encouraged to discuss it with their teammate (the student they shared the computer with). The students were not allowed to discuss their questions with other teams or the teacher-supervisor. Our protocol required that the teachers could only help students with scheduling, computer hardware issues and software operations.[7] In fact, according to our observations, the sessions were so intense that the students were almost always exclusively focused on their computers. There was little communication among the groups or between any of the groups and the teacher-supervisor. The CAL software had enough content and exercise games to cover the math course materials for the entire school year 2011-2012 and the material was sufficient to provide 80 minutes of remedial tutoring per week (two 40-minute sessions).

With both software and hardware ready, we then worked out a detailed CAL curriculum and implementation protocol. The protocol was targeted exclusively at the teacher-supervisors that were responsible for implementing the CAL program in each school. The CAL curriculum was designed to keep pace with the progress of school instruction on a week-by-week basis. This was done so that our CAL sessions provided a timely review and an opportunity to practice the knowledge and skills that were introduced and covered as part of their regular math class. One of the most important jobs of the teacher-supervisor was to make sure the weekly CAL sessions proceeded on a pace that matched the pace of the students' regular math classes. Because this work was clearly beyond the scope of their normal classroom duties, we compensated the teacher-supervisors with a monthly stipend of 100 yuan (approximately 15 USD), an amount roughly equivalent to 15 percent of the wage of a typical rural teacher. All teacher-supervisors of the 36 treatment schools also participated in a two-day mandatory training program.

CAL Control Group (the students in the 36 control schools)

The third-grade and fifth-grade students in the 36 control schools constituted the CAL control group. Students in the control group did not receive any CAL intervention. To avoid any type of the spillover effects of the CAL intervention, the principals, teachers and students (and their parents) of the control schools were not informed of the CAL project. The research team did not visit the control schools except for during the baseline and final evaluation surveys. The students in the control group took their regular math and computer classes at school as usual.

Data Collection

The research group conducted two rounds of surveys in the 72 control and treatment schools. The first-round survey was a baseline survey conducted with all third and fifth graders in the 72 schools in June 2011 at the end of the spring semester and before any implementation of CAL program had begun. The second-round survey was a final evaluation survey conducted at the end of the program in June 2012.

In each round of the survey, the enumeration team visited each school (treatment and control schools) and conducted a two-part survey. In the first part students were given a standardized math test, which gave us our main outcome variable. The math test included 29-31 questions (tests in different grades and rounds included slightly different numbers of questions).[8] All the questions were chosen to not repeat the questions that were contained in the exercises in the CAL software. Students were required to finish tests in each subject in 25 minutes. Time limits were strictly enforced.

In the second part enumerators collected data on the characteristics of students and their families. From this part of the survey we are able to create demographic and socioeconomic variables. The dataset includes measures of each student's *gender*, *age* (measured in years), whether the student is a *boarding student*, has the student *ever repeated a grade*, if the student is the *only child* of his or her family, father's education level (*father has at least junior high school degree*), mother's education level (*mother has at least junior high school degree*) and parental care (*at least one parent lives at home*). To create the indicator of family wealth, we documented the ownership of household appliances to proxy the family asset value. The variable of family wealth equals 1 if the family assets are higher than the median value and 0 otherwise.

To control for Hawthorne effect, unannounced visits were made randomly to all the schools in the sample (including the control schools). Therefore, if the visits caused any changes in the behaviors of the students, teachers or principals, we have no reason to believe that it confounded the treatment effect.

Statistical Methods

We used both unadjusted and adjusted ordinary least squares (OLS) regression analyses to estimate how the academic outcome changed in the treatment group relative to the control group. Our unadjusted analysis regressed the outcome variable (i.e. post-program math test score[9]) on a dummy variable of the treatment (CAL intervention) status. We used adjusted analysis as well to improve statistical efficiency. In all regressions, we corrected for school-level clustering (relaxing the assumption that disturbance terms are independent and identically distributed within schools).

The model we estimated is:

$$y_{isc}=\alpha+\beta*treatment_s+\theta*y_{0isc}+X_{isc}\gamma+\varepsilon_{isc} \quad , \qquad (1)$$

where y_{isc} is the outcome variable after the CAL program for child i in school s and class c, $treatment_s$ is a dummy variable for a student attending a treatment school (equal to one for students in the treatment group and zero otherwise) and ε_{isc} is a random disturbance term clustered at the school level. We also included a set of control variables. Specifically, we controlled for y_{0isc}, the pre-program math test score and Chinese test score for student i in school s and class c, and X_{isc}, a vector of additional control variables. The control variables are expected to only improve the precision of the estimates. The variables in X_{isc} are student and family characteristics (*gender, age, boarding student, ever repeated a grade, only child*, *age of father, age of mother, father has at least junior high school degree, mother has at least junior high school degree, at least one parent lives at home and family wealth*). By construction, the coefficient of the dummy variable $treatment_s$, β is equal to the unconditional difference in the outcome (y_{isc}) between the treatment and control groups over the program period. In other words, β measures how the treatment group changed in the standardized math test score levels after the CAL program relative to the control group. We estimate Equation (1) with control variables (adjusted model) and without control variables (unadjusted model).

The attrition pattern does not differ between the treatment and control groups. The results comparing the attrition rates between the treatment group and the control group show that the attrition rates are not affected by the treatment status (Table 1). In conducting the test, we estimate Equation (1) with the attrition status as the dependent variable and without control variables. As the results show, the attrition rates are not correlated with the treatment status when pooling the third grade and fifth grade students (column 1, row 1). Similarly, when testing attrition rates separately for the third and fifth grade students, no significant difference is found between the treatment and control groups (columns 2 and 3, row 1). All the coefficients are insignificant and close to zero.

We used a set of student characteristics to check the validity of the random assignment. We estimate Equation (1) without control variables by using the baseline characteristics each at a time as the dependent variable. According to our data, we found that for all student characteristics, none of the differences between the treatment and control groups were statistically significant among the samples before attrition (Appendix 2) or among the samples after attrition (Appendix 3). The assignment of treatment is not correlated with any of the student characteristics for the sample students before attrition (Appendix 2, column1). In addition, the differences are almost all small in magnitude. Consistently, none of the student characteristics are significantly different between the treatment and control groups when testing the validity of random assignment separately for the third and the fifth grade students before attrition (Appendix 2, columns 3 and 5). We found the same results on the sample students after attrition (Appendix 3). In other words, our results show that student characteristics are well balanced between the treatment and control groups both before and after attrition.

Results

According to our analysis, the in-school CAL program significantly improved the academic performance of the students in the sample treatment schools (Table 2). The multivariate regression analyses (adjusted and unadjusted) show that by pooling the third grade and fifth grade students, the program impact is estimated to be 0.18 standard deviations (although not significant at the 10 percent level) using the unadjusted model (column 1, row 1). By controlling for student baseline scores and other characteristics, the standard error is largely reduced and the estimate is significant at the 1 percent level. The program impact slightly decreased to 0.16 standard deviations using the adjusted model in Equation 2 (column 2, row 1).

Using only third grade students or only fifth grade students, in the unadjusted model the estimated CAL treatment effects on math test scores are equal to 0.20 standard deviations for the third grade students (Table 3, column 1, row 1), and 0.17 standard deviations for the fifth grade students (Table 3, column 2, row 1). Both of the coefficients of interest in the unadjusted model are not significant (Table 3, columns 1 and 2, row 1).

Next we estimate the treatment impact by including control variables to improve the precision of the estimates. When we add the control variables (using the adjusted model), the more efficient estimates show that the CAL program had a positive and significant impact on the standardized math scores of the students. The estimated treatment effect for third grade students is 0.17 standard deviations (Table 3, column 3, row 1) and is significant at the 10 percent level. The estimated treatment effect for fifth grade students remains at 0.17 standard deviations and is significant at the 5 percent level (Table 3, column 4, row 1).

So how big an effect size is 0.17 standard deviations? According to McEwan (2013) and Schagen and Hoden (2009), educators are often interested in promoting new programs if they have a 0.2 standard deviations effect on test scores. In a majority of the papers that were reviewed by a team (Krishnaratne, White and Carpenter, 2013), the most commonly effect size used for power calculations is 0.2 standard deviations. The implicit assumption is that 0.2 is big enough of an effect to care about. As seen in the paper, the in-school CAL program managed to improve student performance by 0.17 standard deviations. According to a paper by Lai et al. (2014), which calculates an urban-rural academic achievement gap, our CAL program's effect size (0.17) could reduce the urban-rural gap by almost 20%. In another study, we found that a similar rise in standardized test scores were associated with a jump in intra-school district school rankings of up to 5 places (out of around 30 places). These studies, together with the interest policymakers in our study area have shown about CAL (they are now sponsoring an upscaling project), we believe that our effect size of 0.17 standard deviations is sufficiently large to attract the interest of policy makers.

The results testing the heterogeneous effects show that the effect of the CAL treatment varies by boarding status only for the third grade students (Table 4).[10] When we interact the treatment variable with students' boarding status, the coefficient on the interaction term is negative but insignificant on the full sample (column 1, row 2). These results suggest that, on average, boarding students do not seem to benefit differently from the program than the non-boarding students. For the third grade students, the coefficient on the interaction term is negative and significant (column 2, row 2). The results suggest that the program effect on the third grade non-boarding students is 0.3 standard deviations, which is 0.21 standard deviations higher than the impact on the third grade boarding students (column 2, rows 1-2). For the fifth grade students, the results show that the program does not differ by boarding status (column 3, row 2). The coefficient of the interaction term is close to zero and insignificant. Both the boarding and non-boarding students improved by 0.17 standard deviations after the CAL program relative to the control group (column 3, rows 1-2).[11]

Using our data on the computer class activities, we also conducted a test on whether the treatment effect differs for schools where students learn basic computer skills in computer classes and the schools where students do not learn these skills. We included an interaction term between the treatment variable and a variable indicating whether the students learn basic computer operations in the regression that estimates the treatment effect (using equation 1). We can only run such a test among the fifth grade students because there are too few third grade students who have these activities. Our results show that we cannot reject the hypothesis that there is no significant difference between having CAL replace learning activities of basic computer operations and other activities in computer classes. The coefficient on the interaction term is not significant. The result table is available upon request.

In-school Program versus Out-of-school Program

Before we conducted the in-school program, we conducted an out-of-school program as an efficacy trial to test whether CAL could be made to work in rural China. As is detailed in Lai et al. (2013), the program was conducted to investigate the impact that CAL had on the poor. After the success of the efficacy trial, we designed to be incorporated into the regular school day (the current in-school CAL study). In other words, these two programs were designed to have two different study goals.

The in-school and out-of-school CAL programs constitute an interesting comparison. First, as stated above, the two programs are different in the way they were integrated into the school day (either in-school or out-of-school). At the same time, the actual content of the two CAL programs remained the same. Second, the two studies were based on samples from the same student population: they were conducted in the same schools among students who were in the same grades in different school years.[12] The third and fifth grade students in 2010 participated in the first phase and the third and fifth grade students in 2011 participated in the second phase. Therefore, we have decided to not only report the evaluation result of phase two, but also link it with phase one for comparison.

To test the effectiveness of out-of-school CAL (the efficacy trial), we conducted a cluster-RCT in Shaanxi Province during the 2010-11 school year. A total of 5967 students in 72 rural Shaanxi schools were involved in the study. Among the students, 2726 students were boarding students and the other 3074 students were non-boarders. The boarding students constituted the sample for the study; the non-boarders only were used as additional controls. In other words, only the boarding students in the 36 treatment schools received the CAL program.

According to the analysis in Lai et al. (2013), which is replicated in Appendix 4 in this paper, the out-of-school CAL program had a positive and significant effect on the math test scores of students in the treatment schools. Appendix 4 includes the results of the regressions when using Equation (1) with control variables. Overall, as seen in Appendix 4, scores went up by 0.13 standard deviations (column1). The impact on third grade students was 0.18 standard deviations (column 2) and the fifth grade students was 0.10 standard deviations (column 3). Although the result for fifth grade students is not significant, Lai et al. (2013) found significant impact on the poorer students.

By contrasting the in-school and out-of-school programs, we find that both programs improved the performance of most of the students who participated in the CAL treatment programs by non-trivial magnitudes. Specifically, by integrating the CAL program into the course of the regular school day, the program improved performance of third and fifth grade students by 0.17 standard deviations. These estimates are comparable to the out-of-school program effect of 0.18 standard deviations on the third grade students and 0.10 standard deviations on the fifth grade students (for students up to the 70th percentile—using their pretest scores).

Although the in-school program is proved to have successfully improved students' performance, there may have been substitution of the in-school program on the third grade boarding students. Table 4 shows that third grade boarding students benefit less from the in-school program than the non-boarding students by 0.21 standard deviations. One possible explanation is that the substitution effect for the boarding students is larger than the non-boarding students in the third grade. The third grade students were likely to receive teacher-supervised/assisted exercise sessions in computer classes that were replaced by CAL (only two schools had computer competency activities in third grade computer classes). As boarding students were less likely to get any assistance at home, it could be that these computer classes are the only chance they have to get assistance in math learning besides the regular math teaching. As a result, the substitution effect of taking away these exercise sessions is larger for the boarding students than the non-boarding students. This may explain why the out-of-school program works better than the in-school program in helping the third grade boarding students.

In order to test for the different substitution effects by boarding status, we run a regression by including the interaction term of three variables, including the treatment status, boarding status and whether the students were having computer competency activities during computer classes. The results show that boarding students benefit more if the CAL classes replaced computer competency activities instead of the supervised exercise sessions. The point estimate in the difference in treatment effect on boarding students is 0.14 standard deviations, although the estimate is not significant as a result of low power (too few third grade students had computer competency activities).[13][14]

Conclusion

In this paper we present the results from a randomized field experiment of a CAL program in 72 rural public schools in Ankang, Shaanxi. The study involves around 5267 third-grade and fifth-grade students. To evaluate the effectiveness of the program we randomly chose 36 schools from the entire sample as treatment schools and the third and fifth grade students in these schools received the CAL intervention. The remaining 36 schools served as control schools. The main intervention was designed to be a math CAL program held during regular school hours (during a regularly scheduled computer class). The students were offered 40 minutes of shared computer time after school, twice a week. During the sessions students played computer-based games that required them to practice using their knowledge of math and relatively simple problem solving skills. The CAL program was tailored to the regular school math curriculum and was remedial in nature, providing the students with drills and exercises with different levels of difficulty.

Our results indicate that the in-school CAL program significantly improved student academic outcomes. Two 40-minute CAL math sessions per week increased the student standardized math scores by 0.17 standard deviations for third grade students and 0.17 standard deviations for fifth grade students. Although out-of-school programs have typically been considered superior to in-school programs, the gains from this in-school program do not vary much from the overall impact of the out-of-school pilot program reported by Lai et al. (2013).

These results suggest that given the possibility of substitution, the in-school program still improves student learning. The reason that our results differ from the Linden study (2008) is that by integrating the CAL program during a relatively unproductive period of time with low teacher absenteeism, the substitution effect may have been minimized. In order to investigate the substitution effect, we examined the differential impacts of CAL in schools that were teaching computer competencies (and other non-academic material) and those that (sometimes) used the classes as review sessions for math, Chinese and English.[15] Our results show that we cannot reject the hypothesis that there is no significant difference between having CAL replace learning activities of basic computer operations and other activities in computer classes. The coefficient on the interaction term is not significant. This seems to indicate that the conditions in China make in-school programs a viable (and effective) means for introducing CAL to rural students.

There are limitations to our paper. For example, we are interested mostly in the academic outcomes of children (e.g., their math scores) rather than competencies in other areas (e.g., a student's ability to use a mouse or operate Microsoft Office software). We also do realize that our sample covers only one prefecture in southern Shaanxi. Although our study is restricted to one prefecture in Northwest China, we believe that many aspects of our study environment apply to not only our study schools but also schools in other parts of rural China. Rural schools in China are homogeneous in many aspects. For example, almost all schools are public. The Ministry of Education requires that all primary schools cover first through sixth grade. The primary school curriculum used by almost all rural provinces is called "renjiaoban." Renjiaoban is produced and distributed by the centrally-administered People's Education Publishing House, which is part of the Ministry of Education. Teaching credentials are uniform across provinces and salaries of nearly all teachers are paid by the national government (especially in the case of poor rural areas). Although there is potential external validity, we know there are limits to how representative our study areas are.

Given China's current effort to put computer rooms in all schools in poor rural areas of western China, these findings are timely, policy-relevant and immediately actionable. As more computers are installed in rural schools, policymakers and school officials will need to explore various options and ultimately decide which type of CAL program to implement. As the results of our study indicate that both in-school and out-of-school CAL programs can produce positive results in rural China, policymakers and school officials can thus select from a broader range of choices when making their CAL program decisions.[16]

In the end, the final decision on whether to implement out-of-school or in-school programs may be different for different schools and different counties. While out-of-school programs may be more effective in some cases (the out-of-school program benefited the third grade boarding students more than the in-school program), in rural China running an after-school program means that some students (such as non-boarding students with long daily commutes) will be less likely to participate. As such, schools need to balance the potentially greater effectiveness of out-of-school programs with the greater inclusiveness and broader reach of in school programs. Perhaps if school officials are given the flexibility and the proper training, they can decide what degree of integration will give their particular institution the best mix of effectiveness and equity.

In addition to these findings, this study makes another point very clear: as more computer rooms are established in rural schools across western China, it is essential that effective educational-based software is also available and that teachers are trained well in using it.

References

Angrist, J., and V. Lavy. 2002. "New Evidence on Classroom Computers and Pupil Learning." *The Economic Journal,* 112 (October): 735-765.

Banerjee, A., S. Cole, E. Duflo, and L. Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *Quarterly Journal of Economics*. 122 (3): 1235-1264.

Barrow, L., L. Markman, and C. Rouse. 2008. "Technology's Edge: The Educational Benefits of Computer-Aided Instruction." National Bureau of Economic Research working paper No. 14240.

Chaudhury, N., J. Hammer, M. Kremer, K. Muralidharan, and K. F. H. Rogers. 2005. "Teacher Absence in India: A Snapshot." *Journal of the European Economic Association,* 3(2): 658-667.

Cole, M. 1996. "*Cultural psychology: A once and future discipline."* Cambridge, MA: Harvard University Press.

Cristia, J., P. Ibarrarán, S. Cueto, A. Santiago, and E. Severín. 2012. "Technology and Child Development: Evidence from the One Laptop Per Child Program." IDB working paper 304.

Dias, P., A. Freedman, P. Medway, and A. Paré. 1999. "*Worlds Apart: Acting and Writing in Academic and Workplace Contexts."* Mahwah, NJ: Lawrence Erlbaum.

Duflo, E., R. Hanna, and S.P. Ryan. 2010. "Incentives Work: Getting Teachers to Come to School." Mimeo, MIT Economics Department.

Dynarski, M. 2007. "Effectiveness of Reading and Mathematics Software Products: Findings from the 1st Student Cohort." Mathematica Report. http://www.mathematica-mpr.com/publications/redirect_pubsdb.asp?strSite=PDFs/effectread.pdf.

Fuchs, T., and L. Woosmann. 2004. "Computers and Student Learning: Bivariate and Multivariate Evidence on the Availability and Use of Computers at Home and at Schools." *Brussels Economic Review*, 47: 359-585.

Glewwe, P., and M. Kremer. 2006. "Schools, Teachers, and Education Outcomes in Developing Countries." *Handbook on the Economics of Education.* New York, NY: Elsevier, 2: 945-1017.

Glewwe, P., and S. Moulin. 2002. "Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya." BREAD Working Paper, Cambridge, MA.

Glewwe, P., and E. Zitzewitz. 2004. "Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya." *Journal of Development Economics,* 74: 251-268.

Goolsbee, A., and J. Guryan. 2006. "The Impact of Internet Subsidies in Public Schools." *The Review of Economics and Statistics*, 88 (2): 336-347.

Guimarães, L. B. D. M., J. L. D. Ribeiro, M. E. Echeveste, and J. J. De Jacques. 2013. A Study of the Use of the Laptop XO in Brazilian Pilot Schools. *Computers & Education*, 69: 263-273.

Hanushek, E. 1986. "The Economics of Schooling: Production and Efficiency in Public

Schools." *Journal of Economic Literature,* 24: 1141-1177.
Hanushek, E.1995. "Interpreting Recent Research on Schooling in Developing Countries." *World Bank Research Observer,* 10: 227-246.
He, F., L. Linden, and M. MacLeod. 2008. "How to Teach English in India: Testing the Relative Productivity of Instruction Methods within the Pratham English Language Education Program." Working Paper. Columbia University Department of Economics.
Hull, G., and K. Schultz (Eds.) 2002. "*School's Out! Bridging Out-of-school Literacies with Classroom Practice."* New York: Teachers College Press.
Jacob, B., and L. Lefgren. 2004. "The Impact of Teacher Training on Student Achievement Quasi-Experimental Evidence from School Reform Efforts in Chicago." *The Journal of Human Resources,* 39 (1): 50-79.
Kirkpatrick, H., and L. Cuban. 1998. "Computers Make Kids Smarter – Right?" *Technos Quarterly for Education and Technology*, 7 (2): 26-31.
Krueger, A., and C. Rouse. 2004. "Putting Computerized Instruction to the Test: A Randomized Evaluation of a 'Scientifically-based' Reading Program." *Economics of Education Review,* 23: 323-338.
Linden, L. 2008. "Computer or Substitute? The Effect of Technology on Student Achievement in India." Working Paper. Columbia University Department of Economics.
Lai, F., R. Luo, L. Zhang, X. Huang, and S. Rozelle. 2011. "Does Computer-Assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Migrant Schools in Beijing." REAP working paper.
Lai, F., L. Zhang, Q. Qu, X. Hu, Y. Shi, M. Boswell, and S. Rozelle. 2013. "Computer Assisted Learning as Extracurricular Tutor? Evidence from a Randomized Experiment in Rural Boarding Schools in Shaanxi." *Journal of Development Effectiveness,* 5(2): 208-231.
Lai, F., L. Zhang, Q. Qu, X. Hu, Y. Shi, M. Boswell, and S. Rozelle. 2012. "Does Computer-Assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Public Schools in Rural Minority Areas in Qinghai, China." REAP working paper.
Liu, Y. and K. Kumar. 2008. "China and India: Different Educational Paths Toward Prosperity." Policy insight 2 (1): 1–4.
Luo, R., Y. Shi, L. Zhang, C. Liu, S. Rozelle, and B. Sharbono 2009. "Malnutrition in China's Rural Boarding Schools: the Case of Primary Schools in Shaanxi Province." *Asia Pacific Journal of Education* 29 (4): 481-501.
Ministry of Education. 2012. "*12th Five-Year Plan for Integrating Information Technology into Education."* http://www.edu.cn/zong_he_870/20120723/t20120723_813704.shtml
Miettinen, R. 1999. "Transcending Traditional School Learning: Teachers' Work and Networks of Learning." In Y. Engeström, R. Miettinen, & R-L. Punamäki (Eds.),

Perspectives on activity theory (pp. 325–344). Cambridge, UK: Cambridge University Press.

Mo, D., L. Zhang, J. Wang, W. Huang, Y. Shi, M. Boswell, and S. Rozelle. 2013. “The Persistence of Gains in Learning from Computer Assisted Learning: Evidence from a Randomized Experiment in Rural Schools in Shaanxi Province.” REAP working paper.

Nara, H., and M. Noda. 2003. Acts of Reading: Exploring Connections in Pedagogy of Japanese. Honolulu: University of Hawaii Press.

National Bureau of Statistics of China (NBSC). 2011. China Statistical Yearbook. China Statistics Press.

Planty, M., W. M, T. Snyder, S. Provasnik, G. Kena, R. Dinkes, A. Kewal Ramani, and J. Kemp. 2008. *The Condition of Education 2008* (NCES 2008-031). National Center for 50 Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

President’s Committee of Advisors on Science and Technology. 1997. “Panel on Educational Technology.” *Report to the President on the Use of Technology to Strengthen K-12 Education in the United States*, Washington DC: Office of the President.

Underwood, C., M. Welsh, M. Gauvain, and S. Duffy. 2000. “Learning at the Edges: Challenges to the Sustainability of Service-Learning in Higher Education.” *Language and Learning Across the Disciplines,* 4(3): 7–26.

World Bank. 2004. “*World Development Report 2004: Making Services Work for the Poor.”* New York, NY: Oxford University Press.

Yuan, X. 2012. “Strategy for Improving Public Service: Push Forward the Integration of ICT in Rural Areas in China (in Chinese).” http://www.stdaily.com/stdaily/content/2012-02/20/content_429399.htm

Yue, A., Y. Shi, F. Chang, C. Yang, H. Wang, H. Yi, R. Luo, C. Liu, L. Zhang, J. Chu, and S. Rozelle. 2012. “Dormitory Management and Boarding Students in China’s Rural Elementary Schools.” REAP working paper.

Zhang, L., Y. Yang, X. Hu, Q. Hu, F. Lai, Y. Shi, M. Boswell, and S. Rozelle. 2012. “The Roots of Tomorrow’s Digital Divide: Documenting Computer Use and Internet Access in China’s Elementary Schools Today.” REAP working paper.

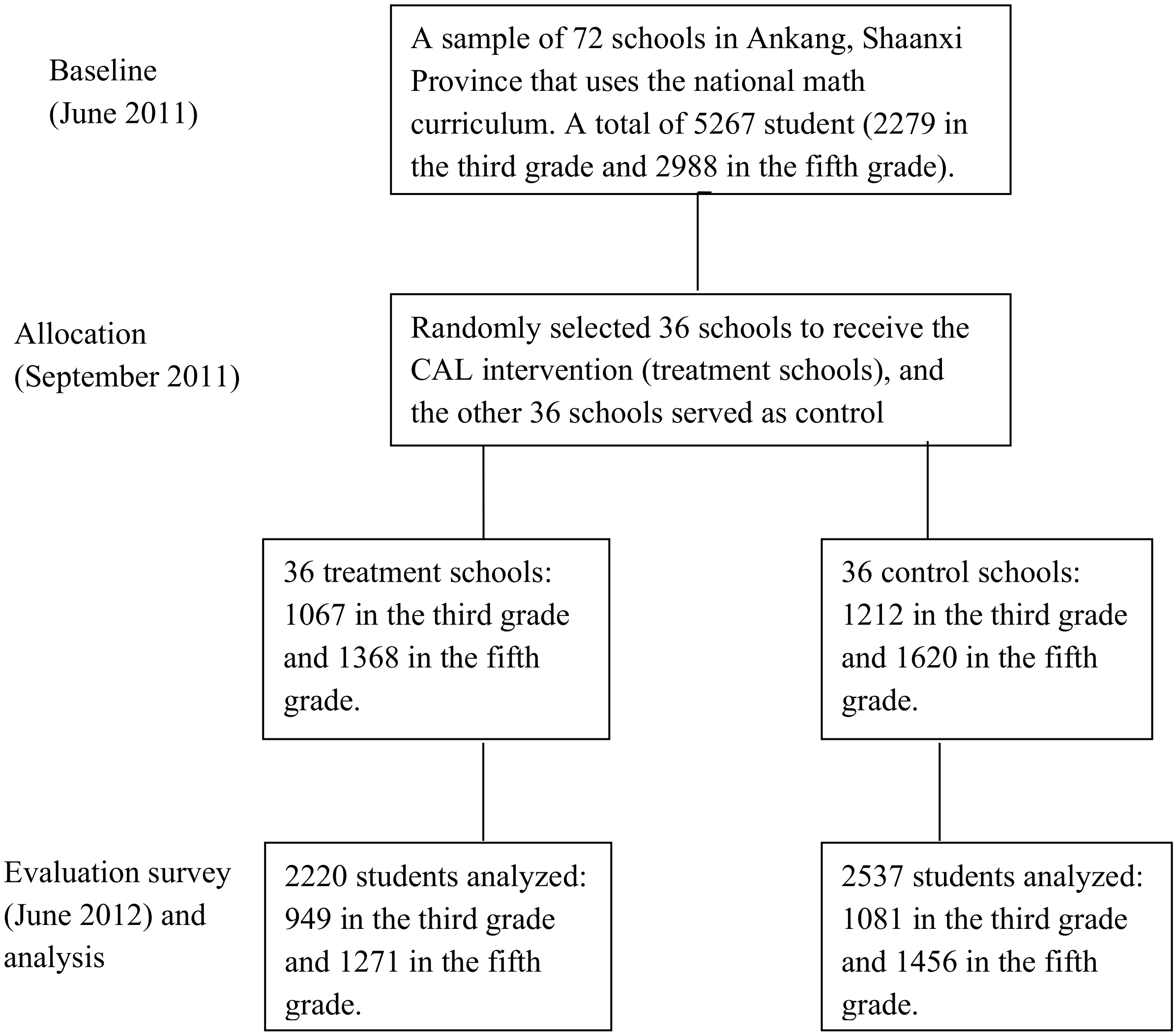

Figure 1: Experiment Profile

Table 1. Comparisons of attrition between the treatment and control students

Dependent variable: attrition (1=students attrited; 0=students remained)				
		All (Third grade & Fifth grade)	Third grade	Fifth grade
		(1)	(2)	(3)
[1]	Treatment (1=treatment group; 0=control group)	-0.02	0.00	-0.03
		(0.02)	(0.03)	(0.02)
[4]	Observations	5,267	2,279	2,988
[5]	R-squared	0.001	0.000	0.003

* significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses clustered at school level.

The test aims to show whether attrition rates are different between the treatment and control groups. The test regresses attrition status on the treatment variable.

Table 2 Ordinary Least Squares estimators of the CAL impact on standardized math test scores on all students (third grade & fifth grade)

Dependent variable: standardized post-CAL math test score (standard deviation)

		(1)	(2)
[1]	Treatment (1=treatment group; 0=control group)	0.18	0.16***
		(0.11)	(0.06)
[2]	Baseline math score (units of standard deviation)[a]		0.39***
			(0.02)
[3]	Baseline Chinese score (units of standard deviation)[b]		0.18***
			(0.02)
[4]	Gender (1=boy; 0=girl)		0.07**
			(0.03)
[5]	Age(years)		-0.03
			(0.02)
[6]	Boarding student (1=yes; 0=no)		0.06
			(0.04)
[7]	Only child (1=yes; 0=no)		-0.13***
			(0.04)
[8]	Ever repeated grade (1=yes; 0=no)		-0.03
			(0.03)
[9]	Age of father (years)		0.00
			(0.00)
[10]	Age of mother (years)		-0.00
			(0.00)
[11]	Father has at least junior high school degree (1=yes; 0=no)		-0.01
			(0.03)
[12]	Mother has at least junior high school degree (1=yes; 0=no)		0.01
			(0.03)
[13]	At least one parent lives at home (1=yes; 0=no)		-0.01
			(0.02)
[14]	Family wealth (1=higher than the median; 0=otherwise)		0.02
			(0.03)
[15]	Observations	4,757	4,757
[16]	R-squared	0.008	0.317

* significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses clustered at school level.

The test aims to show the impact of the in-school CAL treatment on student math test scores of both the third and fifth grade students. The test regresses standardized post-CAL math test scores on the treatment variable and a set of control variables.

[ab] The baseline math/Chinese score is the normalized score on the math/Chinese test that is given to all sample students before the CAL program.

Table 3. Ordinary Least Squares estimators of the CAL impact on standardized math test scores of the third grade and fifth grade students

Dependent variable: standardized post-CAL math test score (standard deviation)

		Third grade	Fifth grade	Third grade	Fifth grade
		(1)	(2)	(3)	(4)
[1]	Treatment (1=treatment group; 0=control group)	0.20	0.17	0.17*	0.17**
		(0.15)	(0.13)	(0.09)	(0.07)
[2]	Baseline math score (units of standard deviation)[a]			0.31***	0.48***
				(0.03)	(0.02)
[3]	Baseline Chinese score (units of standard deviation)[b]			0.17***	0.17***
				(0.04)	(0.02)
[4]	Gender (1=boy; 0=girl)			0.00	0.12***
				(0.04)	(0.03)
[5]	Boarding student (1=yes; 0=no)			-0.06***	-0.08***
				(0.02)	(0.03)
[6]	Age(years)			0.13	0.01
				(0.08)	(0.04)
[7]	Only child (1=yes; 0=no)			-0.14**	-0.10***
				(0.05)	(0.03)
[8]	Ever repeated grade (1=yes; 0=no)			-0.08*	-0.01
				(0.04)	(0.03)
[9]	Age of father (years)			0.01	-0.00
				(0.00)	(0.01)
[10]	Age of mother (years)			-0.00	0.00
				(0.00)	(0.01)
[11]	Father has at least junior high school degree (1=yes; 0=no)			0.01	-0.02
				(0.04)	(0.03)
[12]	Mother has at least junior high school degree (1=yes; 0=no)			0.00	0.02
				(0.05)	(0.03)
[13]	At least one parent lives at home (1=yes; 0=no)			-0.05	0.04
				(0.04)	(0.03)
[14]	Family wealth (1=higher than the median; 0=otherwise)			0.01	0.02
				(0.04)	(0.03)
[15]	Observations	2,030	2,727	2,030	2,727
[16]	R-squared	0.010	0.007	0.299	0.381

* significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses clustered at school level.

The test aims to show the impact of the in-school CAL treatment on student math test scores separately for the third and the fifth grade students. The test regresses standardized post-CAL math test scores on the treatment variable with or without a set of control variables.

[ab] The baseline math/Chinese score is the normalized score on the math/Chinese test that is given to all sample students before the CAL Program.

Table 4. Ordinary Least Squares analysis of the heterogeneous impact of CAL Program on student standardized math test scores of boarding and non-boarding students

Dependent variable: standardized post-CAL math test score (standard deviation)			
		Third grade	Fifth grade
		(1)	(2)
[1]	Treatment (1=treatment group; 0=control group)	0.30***	0.17**
		(0.09)	(0.08)
[2]	Interaction: Treatment * Boarding student	-0.21**	-0.01
		(0.10)	(0.08)
[3]	Boarding student (1=yes; 0=no)	0.17**	0.01
		(0.08)	(0.04)
[4]	Control variables [a]	Yes	Yes
[5]	Observations	2,030	2,727
[6]	R-squared	0.318	0.381

* significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses clustered at school level.

The test aims to show the heterogeneous effects of the in-school CAL treatment by student boarding status. The test regresses standardized post-CAL math test scores on the treatment variable, the interaction term between boarding and treatment status, boarding status and a set of control variables.

[a] Control variables include all the variables in Appendix 1.

Appendix 1. Descriptive statistics of baseline characteristics of the treatment group and the control group of the third grade and fifth grade students after attrition

		Students after attrition			
		Treatment group		Control group	
		Mean	Standard deviation	Mean	Standard deviation
[1]	Baseline math score (units of standard deviation)[a]	0.00	1.02	0.00	1.00
[2]	Baseline Chinese score (units of standard deviation)[b]	0.00	1.00	0.00	1.00
[3]	Gender (1=boy; 0=girl)	0.54	0.50	0.54	0.50
[4]	Age(years)	9.74	1.27	9.76	1.21
[5]	Boarding student (1=yes; 0=no)	0.37	0.48	0.35	0.48
[6]	Ever repeated grade (1=yes; 0=no)	0.29	0.46	0.28	0.45
[7]	Only child (1=yes; 0=no)	0.29	0.45	0.28	0.45
[8]	Father has at least junior high school degree (1=yes; 0=no)	0.55	0.50	0.56	0.50
[9]	Mother has at least junior high school degree (1=yes; 0=no)	0.40	0.49	0.41	0.49
[10]	At least one parent lives at home (1=yes; 0=no)	0.42	0.49	0.43	0.49
[11]	Family wealth (1=higher than the median; 0=otherwise)	0.50	0.50	0.50	0.50
[12]	Observations	2220		2537	

[ab] The baseline math/Chinese score is the normalized score on the math/Chinese test that is given to all sample students before the CAL Program.

Appendix 2. Comparison of characteristics between the treatment group and the control group of the third grade and fifth grade students before attrition

		Students before attrition					
		Third grade & fifth grade		Third grade only		Fifth grade only	
		Coefficient	Standard error	Coefficient	Standard error	Coefficient	Standard error
[1]	Baseline math score (units of standard deviation)[a]	0.01	0.03	0.02	0.03	0.01	0.04
[2]	Baseline Chinese score (units of standard deviation)[b]	0.02	0.03	0.01	0.03	0.02	0.03
[3]	Gender (1=boy; 0=girl)	0.01	0.01	0.00	0.02	0.01	0.02
[4]	Age (years)	-0.01	0.01	-0.03	0.02	0.01	0.03
[5]	Boarding student (1=yes; 0=no)	0.02	0.07	0.01	0.09	0.02	0.07
[6]	Ever repeated grade (1=yes; 0=no)	0.01	0.03	-0.02	0.05	0.03	0.04
[7]	Only child (1=yes; 0=no)	0.01	0.04	0.03	0.05	0.00	0.04
[8]	Father has at least junior high school degree (1=yes; 0=no)	0.00	0.03	0.01	0.03	-0.01	0.04
[9]	Mother has at least junior high school degree (1=yes; 0=no)	-0.01	0.03	-0.01	0.03	-0.02	0.04
[10]	At least one parent lives at home (1=yes; 0=no)	-0.01	0.03	0.00	0.03	-0.03	0.03
[11]	Family wealth (1=higher than the median; 0=otherwise)	0.00	0.04	0.00	0.04	0.00	0.04
[12]	Observations	5267		2279		2988	

The test aims to show whether the samples are well balanced in the treatment and control groups before attrition for the third and fifth grade students (combined and separately). These tests regress the student and family characteristics on the treatment status one at a time.

[ab] The baseline math/Chinese score is the normalized score on the math/Chinese test that is given to all sample students before the CAL program.

Appendix 3. Comparison of characteristics between the treatment group and the control group of the third grade and fifth grade students after attrition

		Students after attrition					
		Third grade & fifth grade		Third grade only		Fifth grade only	
		Coefficient	Standard error	Coefficient	Standard error	Coefficient	Standard error
[1]	Baseline math score (units of standard deviation)[a]	0.01	0.03	0.02	0.03	0.01	0.04
[2]	Baseline Chinese score (units of standard deviation)[b]	0.02	0.03	0.01	0.03	0.02	0.03
[3]	Gender (1=boy; 0=girl)	0.00	0.01	-0.01	0.02	0.01	0.02
[4]	Age (years)	0.00	0.01	-0.03	0.02	0.01	0.03
[5]	Boarding student (1=yes; 0=no)	0.02	0.08	0.03	0.09	0.01	0.07
[6]	Ever repeated grade (1=yes; 0=no)	0.01	0.04	-0.02	0.06	0.04	0.04
[7]	Only child (1=yes; 0=no)	0.01	0.05	0.03	0.05	0.00	0.05
[8]	Father has at least junior high school degree (1=yes; 0=no)	-0.01	0.03	0.01	0.03	-0.02	0.04
[9]	Mother has at least junior high school degree (1=yes; 0=no)	-0.01	0.03	0.00	0.03	-0.02	0.04
[10]	At least one parent lives at home (1=yes; 0=no)	-0.01	0.03	0.01	0.04	-0.02	0.03
[11]	Family wealth (1=higher than the median; 0=otherwise)	-0.01	0.04	-0.01	0.04	-0.01	0.04
[12]	Observations	4757		2030		2727	

The test aims to show whether the samples are well balanced in the treatment and control groups after attrition for the third and fifth grade students (combined and separately). These tests regress the student and family characteristics on the treatment status one at a time.

[ab] The baseline math/Chinese score is the normalized score on the math/Chinese test that is given to all sample students before the CAL program.

Appendix 4. Ordinary Least Squares analysis of the out-of-school CAL Program on student standardized math test scores of boarding students of the 3rd and 5th grade students who participated in the CAL Program during March and June, 2011

Dependent variable: standardized post-CAL math test score (standard deviation)

		All (Third grade & Fifth grade)	Third grade	Fifth grade
		(1)	(2)	(3)
[1]	Treatment (1=treatment group; 0=control group)	0.13**	0.18**	0.10
		(0.06)	(0.09)	(0.07)
[3]	Control variables	Yes	Yes	Yes
[4]	Observations	2,426	1,038	1,388
[5]	R-squared	0.427	0.416	0.453

* significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses clustered at school level.

This is a replication of the analysis in Lai et al. (2013). The test aims to show the impact of the out-of-school CAL treatment on student math test scores. The test regresses standardized post-CAL math test scores on the treatment variable and a set of control variables that are listed in Appendix 1.

Appendix 5. Ordinary Least Squares analysis of the in-school CAL Program on student standardized Chinese test scores

Dependent variable: standardized post-CAL Chinese test score (standard deviation)				
		All (Third grade & Fifth grade)	Third grade	Fifth grade
		(1)	(2)	(3)
[1]	Treatment (1=treatment group; 0=control group)	0.05	0.03	0.07
		(0.06)	(0.06)	(0.10)
[3]	Control variables [a]	Yes	Yes	Yes
[4]	Observations	2,426	1,038	1,388
[5]	R-squared	0.345	0.430	0.319

Robust standard errors in parentheses clustered at school level.
The test aims to show the spillover effect of the in-school CAL treatment on standardized post-CAL Chinese test scores. The test regresses standardized post-CAL Chinese test scores on the treatment variable and a set of control variables.
[a] Control variables include all the variables in Appendix 3.

Appendix 6. CAL hardware and software

The intervention team spent considerable time to prepare the necessary hardware, software, CAL curriculum and program implementation protocol in a way that would both facilitate smooth implementation of the CAL program and avoid confounding influences that might bias our results. As the first step, to meet the hardware requirements of the CAL program, we acquired (by way of donation from Dell, Inc.) 640 brand new identical desktop computers and installed our CAL software package on these desktops. We then removed all pre-installed software that would not be used during the CAL intervention (such as Windows built-in games and Microsoft Office) and disabled the Internet and USB functions on all of the computers. In this way, not only could we prevent students or teachers from using the program computers for other purposes that might affect the operation of the regular CAL program but we could also avoid the interruptions that might otherwise be caused by accidental deletion of the CAL software or the introduction of viruses. “Sealing the computers” also ensured the quality of our evaluation of the program effects without capturing any other confounding influences (spillovers) if students used the computers to gain knowledge by having access to other sources of information such as the Internet. It also prevented teachers and students from the control schools from copying our CAL software onto their computers.

The educational program contained in the computer assisted learning software has two parts. The first piece of software was a commercial, game-based math-learning software program that was obtained via donation. We adopted this package because it was uniquely suited to the CAL program. The software provided remedial tutoring material (both animated reviews and remedial questions) in math for the third and fifth grade students following the national uniform math curriculum. The designers of the program also developed their software so it could be used in conjunction with the material that students were learning in their math class on a week by week basis.

We also developed the second piece of software by ourselves. Our software package (henceforth, the CAL software) was developed to provide the students with a large number of practice questions. Students answered the questions in game-based exercises. In choosing the math questions to include in the CAL software, we consulted experienced elementary school math teachers in both public schools in cities and rural areas, as well as expert committee members of the Center for Examination of Beijing, an institute that designs city-wide uniform tests for elementary schools in Beijing. With their direction and assistance, we chose questions for the CAL software from several commercially available books of practice questions. By combining the commercial software and the CAL software, we had enough content and exercise games to provide 80 minutes of remedial tutoring (two weekly sessions of 40 minutes each) that cover the math curriculum for the spring 2011 semester.

We also developed the second piece of software by ourselves. Our software package (henceforth, the CAL software) was developed to provide the students with a large number of practice questions. Students answered the questions in game-based exercises. In choosing the math questions to include in the CAL software, we consulted

experienced elementary school math teachers in both public schools in cities and rural areas, as well as expert committee members of the Center for Examination of Beijing, an institute that designs city-wide uniform tests for elementary schools in Beijing. With their direction and assistance, we chose questions for the CAL software from several commercially available books of practice questions. By combining the commercial software and the CAL software, we had enough content and exercise games to provide 80 minutes of remedial tutoring (two weekly sessions of 40 minutes each) that cover the math curriculum for the spring 2011 semester.

---

[1] The study underwent and successfully passed ethical review by Stanford University's Internal Review

[2] In terms of educational achievement, Shaanxi Province is at about the national average. However, there is huge inequality within the province (NBSC, 2011). For example, in relatively rich areas such as Guanzhong (in the central part of the province), 14.4% of its population received a college education (higher than the national average of 8.9%). In contrast, in Ankang Prefecture only 4.8% of the population holds a college degree.

[3] We only included wanxiao (or "complete schools") with six full grades in our sample because the program requires that third grade and fifth grade students stay in the same school during the program period (one year and a half). In rural China, there are other elementary schools with only two, three and four grades. These are often small schools (several students per grade) in remote rural villages. In Chinese these are called *jiaoxuedian,* or "teaching point schools". The schools that were not complete schools could not be included in the program because students often transfer to other schools when they reach the third or fourth grades. Teaching point schools also are being shut down and merged into larger complete schools. In either case, it would be impossible for students to continue to attend the CAL sessions. It would also be hard for us to follow the students. Therefore, non-wanxiao schools were excluded from the sampling frame.

[4] In selecting the teacher-supervisors, we were guided by two principles. First: we wanted to choose the teacher-supervisor rather than the school principal. We also did not want to select a teacher-supervisor that was also a math teacher. With these principles in mind, we excluded from our selection the math teachers or homeroom teachers of the third- and fifth-grade students. We then created a list of teachers that were available. We then randomly chose the teacher.

[5] In terms of teacher training, all teacher-supervisors of the 36 treatment schools participated in a two-day mandatory training program. During the training, the teachers were introduced to our program protocol and

the two pieces of software. The teachers also underwent hands on session where they practiced with the software and asked questions. At the end of the training session, randomly selected teachers gave mock classes to all the other teachers who pretended to be students.

[6] Both the third and the fifth grade CAL software packages consisted of two individual pieces of software. The first piece of software was a commercial, game-based math-learning software program that we obtained via donation. The software provided remedial tutoring material (both animated reviews and remedial questions) in math for the third and fifth grade students in keeping with the national uniform math curriculum. We developed the second piece of software ourselves. The package (henceforth, the CAL software) was designed to provide the students with a large number of remedial questions.

[7] The students were not allowed to discuss math questions with the supervising teacher because the goal of the study is to test whether a CAL program can improve learning of the underserved students in rural schools. We are interested in knowing whether the program can benefit students in the poorest schools with little teaching resources. Therefore, we would like to isolate the program impact from teacher instruction. In other words, teachers were not allowed to intervene in the classes to affect the program impact. In fact, this is policy relevant given a scenario in which the CAL sessions were run in-school during computer class sessions. The computer teacher would not be an expert in the field and would likely be busy managing the technology and curriculum and not focused on teaching students the material that other teachers were supposed to be teaching. Likewise, the students were not allowed to discuss with other teams (students using a different computer) to limit the influence of student interaction on program impact. It also makes it easier for teachers to manage the classes without having to organize the group discussion or other activities. According to our observation, students typically had no time to discuss with each other because the sessions were so intense that the students were almost always exclusively focused on their computers.

[8] The test questions for the standardized math exam were chosen from the TIMSS test data bank. Drafts of the tests were screened by a set of rural elementary teachers in Shaanxi province. We then rigorously tested the questions in a pilot survey. We then made adjustment to the test by eliminating the questions that were too difficult (almost no one got them right) and the questions that were too easy (almost everyone got them right).

[9] The standardized test scores are normalized using the distribution of test scores of the control group students within the same grade and on the same subject.

[10] This is important to show because the original analysis for the out-of-school treatment effects for CAL was conducted for boarding school students only. If we show that the effects are the same for boarding and non-boarding students (as we do), then the analysis can really focus on differences between in-school effects (as reported from this study) and out-of-school effects (as shown in Lai et al., 2011).

[11] We have also conducted analysis of heterogeneous effects by student baseline math score, student gender, family wealth and their starting grade. The questions that the tests are intended to address are whether poorer performing students benefit more from the program than better performing students, whether boys or girls benefit more from the program, whether poorer or richer students improve more after the program and whether starting grade (third or fifth grade) makes a difference in how much student learning can be improved. The results show that none of the tests detect any significant difference in program impact among the subgroups. By following the Bonferroni approach to adjust multiple hypotheses, we divide the

significance level of all the correlated outcomes of heterogeneous effects, 0.1, by the number of hypotheses we tested (i.e. 4 different types of heterogeneous effects of CAL program). By doing this, we get the adjusted p-values for each individual null hypothesis of heterogeneous effects: 0.1/4=0.025. Since none of the heterogeneous effects are significant at 0.1 level, they do not meet the 0.025 adjusted significance requirement, either. In other words, with or without adjusting for multiple hypotheses testing, we cannot reject the null hypotheses that there are no heterogeneous effects between the poorer and better performing students, between girls and boys, between the richer and poorer students, and between the third grade and the fifth grade students.

[12] Although no re-randomization was done to re-assign treatment and control schools in the in-school program, we did conduct a balance test before the start of the program to make sure that the students in the two groups were balanced. As shown in Appendix 2, the key variables of the treatment and control groups are balanced at the baseline.

[13] The result table is available upon request.

[14] We also tested whether the program had any crowding out effect on Chinese learning. Based on the regression results using Equation (1), the out-of-school program does not seem to have crowded out student learning in Chinese (Appendix 5). The coefficient of the treatment variable is not significant for either the whole sample (third and fifth grade) or each grade separately. The magnitudes of the coefficients are small and positive.

[15] Using our data on the computer class activities, we conducted a test on whether the treatment effect differs for schools where students learn basic computer skills in computer classes and the schools where students do not learn these skills. We included an interaction term between the treatment variable and a variable indicating whether the students learn basic computer operations in the regression that estimates the treatment effect (using equation 2). We can only run such a test among the fifth grade students because there are too few third grade students who have these activities. The result table is available upon request.

[16] Cost-effectiveness analysis suggests that the program has low cost per unit of improvement in student learning. From the perspective of China's policymakers considering to upscale the program, computer hardware itself is already a sunk cost given that the government is installing computer labs in every rural elementary school as part of its Twelfth Five Year Plan. The marginal costs that are needed to execute the program include teacher training, administration costs and allowance for CAL teacher-supervisors. Using the method suggested by Dhaliwal et al. (2012), we calculate the total cost of the program in our project area to be 9,439 USD (in 2011, the project year) and 10,035 USD (in 2014, after taking inflation into account). We then divide the total cost by total impact (total impact=average program effect multiplied by the total number of students attending CAL sessions): 10,035 USD/(0.17 SD * 2435 students)=24.2 USD/SD. The cost-effectiveness of our program is comparable to the CAL program conducted in India. According to the estimates provided by Banerjee et al. (2007), the CAL program in India costs 21.4 USD/SD (in 2002) and 28.2 USD/SD (in 2014)—also excluding the costs of computers.